Unreliable Inferences about Unobserved Processes

A Critique of Partial Observability Models*

Carlisle Rainey
Department of Political Science
University at Buffalo (SUNY)
Buffalo, NY 14260
carlislerainey.com
rcrainey@buffalo.edu

Robert A. Jackson
Department of Political Science
Florida State University
Tallahassee, FL 32306
rjackson@fsu.edu

Abstract

Methodologists in political science have begun to advocate the partial observability model (aka the split population model) as an approach that enables researchers to draw accurate inferences about outcome variables that are only partially observable. Although the partial observability model is theoretically-driven and statistically sound, we suggest that it can lead to large inferential errors. We use simulations to show that seemingly innocuous model specification errors that have little impact under full observability can lead to large biases, including sign errors, under partial observability. To illustrate empirically the misleading inferences that can emerge, we estimate a partial observability model relying on self-reported voter turnout data, arguing that similar, though distinct, processes lead to turning out to vote and misreporting voting if one abstains. We show that using a partial observability model to model the unobserved process leads to worse inferences than ignoring the problem.

...spurious verisimilitude and spurious detailing [are] the two greatest vices of the statistician. Once you have a formula and an electronic computer, there is an awful temptation to squeeze the lemon until it is dry and to present a picture of the future which through its very precision and verisimilitude carries conviction. Yet a man who uses an imaginary map, thinking it to be a true one, is likely to be worse off than someone with no map at all; for he will fail to inquire whenever he can, to observe every detail on his way, and to search continuously with all his senses and all his intelligence for indications of where he should go.

E.F. Schumacher, *Small is Beautiful*

* We thank Will Moore, participants at the 2012 Southern Political Science Association Annual Conference, and participants at the 2012 Midwest Political Science Association Annual Conference for valuable comments on previous versions of the manuscript. We thank John D. Cook for pointing us to the quote above, found on pp. 248-249 of the 1989 edition of E.F. Schumacher's book *Small is Beautiful*. All code and data necessary to replicate the simulations and empirical analysis are available at github.com/carlislerainey/Unreliable.

Introduction

Political scientists often face situations in which two (or more) distinct processes lead to different unobservable binary outcomes that can only be observed jointly. In some applications, researchers observe an event if either unobservable event occurs. In other cases, researchers only observe an event if both occur. Braumoeller (2003) provides many examples of established literature theorizing such relationships. In some cases, a single variable influences both unobservable events in a similar fashion—or perhaps this variable has a positive relationship with one and a negative relationship with the other. Recent work in political methodology suggests that scholars can draw compelling inferences about the effect of this variable on each unobserved outcome via the partial observability model (aka the split population model) as long as the researcher can identify at least one other variable that influences only one of the processes (Przeworski and Vreeland 2002).[1]

Research employing the partial observability model has already weighed in on political scientists' understanding of several major outcomes and processes (e.g., conflict and trade, IMF agreements, and regulatory compliance), and the model would appear to hold out promise for investigating numerous other subjects — including deterrence and treaty compliance and attitudes and behaviors that are subject to social desirability bias when measured via a survey report — for which arriving at convincing answers to central questions has proven to be elusive for empirical researchers.[2] In some specific applications, the partial observability model may

[1] Poirier (1980) presents the partial observability model, and Abowd and Farber (1982) provide an early economics application.

[2] Feinstein (1990) advocates the approach for regulatory compliance studies and, more generally, suggests the model's usefulness for a wide range of policy studies. The most well-known political science applications of the partial observability model are the studies of IMF agreements and of trade and conflict that we discuss more fully below (Przeworski and Vreeland 2000 and 2002, Vreeland 2003, Stone 2008, and Xiang 2010). In addition, Beger et al. (2011) recommend that researchers consider the partial observability model when assessing behaviors or attitudes that are subject to social desirability bias when measured via survey reports, as well as when investigating deterrence and treaty compliance.

indeed be the best approach.[3] However, we believe that existing research likely has made overly strong claims based on results from the model and that future researchers should exercise due caution and invoke numerous robustness checks before employing the model. Our reservations are based on 1) findings from simulations and 2) an empirical investigation that reveals marked incongruence between the estimates from a partial observability model and those from a model that relies on (approximately) fully-observed data for the same respondents.

The Partial Observability Model

Some Recent Applications

Literature on conflict suggests that trade between nation states should have a pacifying effect, but only among "politically relevant" dyads. Indeed, the literature suggests that "politically irrelevant" states have no opportunity for conflict and thus trade should have no effect among these states. Yet trade is at least indirectly related to many concepts of relevancy, leading Xiang (2010) to suggest that political scientists use a partial observability model to model the unobserved concept of relevance and then to model the probability of conflict conditional on relevance.

$$p(conflict \mid relevance) = f(X\beta)$$

$$p(relevance) = f(Z\gamma)$$

$$p(conflict) = p(conflict \mid relevance) \times p(relevance) = f(X\beta) \times f(Z\gamma)$$

Using this approach, Xiang suggests that trade has a small pacifying effect on the probability of conflict given relevance, but a huge positive effect on relevance. Combining trade's effects on relevance and conflict conditional on relevance, the model suggests that trade initially has a large positive effect on the probability of conflict, but that the effect becomes negative as trade

[3] Collecting data that directly measure the outcome of interest remains the holy grail, of course, but doing so is not always plausible or even possible.

increases.

Przeworski and Vreeland (2002) similarly argue that a partial observability model can be used to parse the effects of several explanatory variables on two different actors whose decisions are only jointly observed. The authors note that in the case of an agreement between two actors, both actors decide whether to enter. However, an agreement is observed only if both actors decide to enter. If only one actor decides to enter, or neither actor decides to enter, then no agreement is observed. Specifically, Przeworski and Vreeland (2002; see also Przeworski and Vreeland 2000, Vreeland 2003, and Stone 2008) are interested in how various factors influence whether the IMF and a national government reach an agreement — e.g., does a surplus in a nation's budget impact both the IMF's decision to enter and the national government's decision to enter. They find that as a budget surplus increases, a government becomes less likely to enter an IMF agreement,[4] but the IMF becomes more likely.

Cautionary Notes

These recent political science applications are sanguine about employing the partial observability model. However, Meng and Schmidt (1985) counseled economists early on regarding some of the costs of partial observability, arguing that standard errors are much larger when the outcome of interest is only partially observed. They write, "we would not be surprised to find, in a typical application, *t*-ratios to be from two to four times as large under full observability as under partial observability" (Meng and Schmidt 1985, p.). Yet standard errors (by design) only reflect the uncertainty due to sampling error. Other sources of error, such as measurement error, missing data, and specification error, create additional uncertainty.

[4] Stone (2008) incorporates a partial observability model to study participation in an IMF program using data from a later time period. However, he reports an opposite effect for an increase in budget surplus regarding a government's decision to participate.

Although innocuous specification errors (particularly of the functional form variety[5]) do not usually introduce much additional uncertainty, we argue that this is not the case with partial observability models. Furthermore, we argue that an even greater cost of partial observability models is the substantial *bias* introduced by small specification errors, which is not reflected in the standard errors. We use simulations to show that misspecifications, which lead to almost no bias under full observability, lead to large biases under partial observability. We then examine a substantively motivated and theoretically interesting application of the partial observability model — self-reported voter turnout in a U.S. presidential election — in which the correct (full observation) inferences are approximately known via a standard binary logit model relying on *validated* voter turnout data for the same respondents. We find that the inferences from the partial observability model are often substantially off the mark, leading to incorrect substantive conclusions.

Study Assumptions

Assume a situation in which a potential dichotomous outcome of interest $\mathrm{d}_{\mathrm{main}}$ cannot be directly observed. Instead, it can only be observed jointly with another dichotomous outcome $d_{nuisance}$.[6] The researcher only observes the outcome variable y_{obs}, which, depending on the application, might equal one if both d_{main} and $d_{nuisance}$ equal one or if either d_{main} or $d_{nuisance}$ equals one. For simplicity, we consider the situation in which y_{obs} equals one if either $d_{main} = 1$ or $d_{nuisance} = 1$. That is, $P(y_{obs}) = P(d_{main}) + P(d_{nuisance}) - P(d_{main}) \times P(d_{nuisance})$. However, all of our findings and suggestions generalize to the other situation as

[5] For example, a functional form error might be including x in the model when the actual specification is *log(x)*.

[6] For clarity, we imagine a situation in which one outcome is of interest and the other is largely nuisance. However, our ideas generalize to a situation in which the researcher's interest extends to both outcomes.

well.[7]

To model the outcome variable y_{obs}, we simply recognize that it is determined jointly by $d_{main}$and $d_{nuisance}$ and assume that a standard logit model relates the matrix of covariates X to $P(d_{main})$, so that $P(d_{main}) = logit^{-1}(X\beta)$. Similarly, assume that the same model relates the matrix of covariates Z to $P(d_{nuisance})$, so that $P(d_{nuisance}) = logit^{-1}(Z\gamma)$. Then, rewriting the equation for $P(y)$ above, we obtain

$$P(y_{obs}) = logit^{-1}(X\beta) \ + logit^{-1}(Z\gamma) + logit^{-1}(X\beta)\times logit^{-1}(Z\gamma).$$

Using this form, it is straightforward to find the log-likelihood function for β and γ, although the maximization is not trivial.[8]

Simulations

To get a sense of how the partial observability model performs under misspecification, we conduct a simulation study, considering a situation in which the researcher does not know the exact model specification, but makes standard modeling assumptions.[9] Although the researcher does not know the exact functional form relating the explanatory variables to the unobserved

[7] In fact, there is an exact mathematical relationship between the two. [I can show the exact relationship, but Google documents doesn't allow math in the footnotes.]

[8] In Appendix II, we provide the log-likelihood function. However, we find that standard "hill-climbing" optimization routines find local maxima. To avoid this, we tried an algorithm suggested by Mebane and Sekhon (1998, 2011) that relies on a combination of genetic adaptation and hill-climbing to efficiently locate the maximum. This approach consistently located the global maximum in several trial examples, but only for a very large population (i.e., computationally prohibitive) of agents. We found that simply starting a standard hill-climbing algorithm at several sets of random starting values consistently located the global maximum. We use ten sets of starting values in the simulation studies and 200 in the empirical illustration. Additionally, partial observability models are prone to problems with separation (i.e., a coefficient with the value of infinity maximizing the likelihood function). To avoid this problem, we use a weakly informative Cauchy prior with scale parameter 2.5, as suggested by Gelman et al. (2008). In general, this prior tends to improve the inferences from the model.

[9] Usually, methodologists encourage applied researchers to make these assumptions for mathematical and computational convenience. For example, suspecting a monotonic relationship between the explanatory variables and the outcome, the researcher usually uses a simple linear specification along with a link function to map the parameters of the distribution to the real line. Our results show that this default choice works well in some applications, even when the functional form is not correctly specified (see our simulations for full observability models, especially the left panel of Figure 4, and Berry, DeMeritt, and Esarey 2014). However, this default choice can have important implications for the inferences from other models, such as the partial observability model, as we argue.

outcomes, she does know that only x influences d_{main} and that both x and z influence $d_{nuisance}$. Suppose the following relationships:

$$P(d^{main}) = f(x), \text{ where } f \text{ is a monotonic function}$$

$$P(d^{nuisance} = 1) = g(x,\ z), \text{ where } g \text{ is a monotonic function of both } x \text{ and } z$$

$$P(y) = f(x) = \begin{cases} 1, & d^{main} = 1 \text{ or } d^{nuisance} = 1 \\ 0, & \text{otherwise} \end{cases}$$

This is less messy than many applied problems, which almost always involve many more variables as well as uncertainty about which variables belong in each equation. In our hypothetical situation, the researcher knows that the relationships are monotonic and knows exactly which variables belong in each equation, so the misspecification, while always present, is slight. Further, the models are relatively simple—only one variable in the main equation and two variables in the nuisance equation. To generate each relationship, we do the following:

1. Assume that the true relationship between $P(d_{main})$ and x is given by $P(d_{main}) = \beta_{cons} + \beta_x x + \beta_{x^2} x^2$. Assume that the true relationship between $P(d_{nuisance})$ and x and z is given by $P(d_{nuisance}) = \gamma_{cons} + \gamma_x x + \gamma_z z + \gamma_{xz} xz + \gamma_{x^2} + \gamma_{z^2} z^2$.
2. Simulate the parameter vectors β and γ from a uniform distribution that ranges from -1 to 1. Repeat until each of the following conditions are met:
    a. Both $P(d_{main})$ and $P(d_{nuisance})$ are bounded between zero and one across the ranges of x and z.
    b. The relationship between $P(d_{main})$ and x is monotonic across the range of x.
    c. The relationship between $P(d_{nuisance})$ and x and z is monotonic (in both x and z) across the ranges of x and z.

Using the procedure above, we simulate 500 data sets with the following steps:

1. Simulate a hypothetical relationship using the steps above.

2. Simulate the explanatory variable x by taking 100,000 draws from a uniform distribution from zero to one. Simulate z similarly.
3. Using the simulated β from Step 1, calculate $P(d_{main}|\ x)$. Simulate $d_{main}$by taking 100,000 draws from a Bernoulli distribution with the success probability set to $P(d_{main}|\ x)$.
4. Similarly, using the simulated γ from Step 1, calculate $P(d_{nuisance}|\ x, z)$. Simulate $d_{nuisance}$by taking 100,000 draws from a Bernoulli distribution with the success probability set to $P(d_{nuisance}|\ x, z)$.
5. Calculate $P(y_{obs})$ using the identity $P(y_{obs}) = P(d_{main}) + P(d_{nuisance}) - P(d_{main}) \times P(d_{nuisance})$. Simulate y_{obs} by taking 100,000 draws from a Bernoulli distribution with the success probability set to $P(y_{obs})$.
6. Calculate the full observability estimate using the following steps:
    a. Estimate the usual logistic regression model $P(d_{main}) = \Delta^{-1}(b_{cons} + b_x x)$.
    b. Use the large-sample estimates b_{cons} and b_x to calculate the predicted probability that $d_{main} = 1$ when x is set to a low value (25th percentile or 0.25) and when x is set to a high value (75th percentile or 0.75). Subtract the latter from the former to obtain an estimate of the effect of changing x from 0.25 to 0.75 on $P(d_{main} = 1)$. Store this estimate.
7. Calculate the full observability estimate using the following steps:
    a. Estimate the partial observability logistic regression model $P(y_{obs}) = \Delta^{-1}(b_{cons} + b_x x) + \Delta^{-1}(g_{cons} + g_x x + g_z z)\times(1 - \Delta^{-1}(b_{cons} + b_x x))$.
    b. Use the large-sample estimates b_{cons} and b_x to calculate the predicted probability that $d_{main} = 1$ when x is set to a low value (25th percentile or 0.25) and when x

is set to a high value (75th percentile or 0.75). Subtract the latter from the former to obtain an estimate of the effect of changing x from 0.25 to 0.75 on $P(d_{main} = 1)$. Store this estimate.

8. Calculate the true effect using the parameters of the hypothetical relationship generated in Step 1 and the true functional form. Store this estimate.

This process yields 500 different data sets and estimates that we can use to assess the size of the large-sample bias of logistic regression models under full and partial observability. They illustrate that misspecification introduces a great deal more uncertainty into the estimations under partial observability than under full observability. Indeed, for the true models we consider, the usual (full observability) logistic regression model is quite robust to misspecification (Berry, DeMeritt, and Esarey 2014). However, the partial observability model is not. To get a sense of how the model is performing, we consider two estimands (potential quantities of interest). First, we examine how well the full and partial observability models estimate the probability of the event of key interest (i.e. $P(d_{main} = 1)$) when the explanatory variable takes on a central value ($x = 0.5$ in our simulations). Second, we use both approaches to estimate how this probability changes as x moves from its 25th to 75th percentile (from 0.25 to 0.75 in our example). Although the estimators are equally good in the theoretical sense that they are both asymptotically consistent but biased in small samples, we show that the partial observability model is much less robust to seemingly innocuous specification errors. Indeed, specification errors that have almost no impact on the inferences from a full observability logistic regression model lead to large biases in a partial observability model.

Consider first Figure 1, which shows how well the full and partial observability models estimate the occurrence probability of the event of interest when the explanatory variable takes on a central value. The figure shows the histogram and density for the large-sample bias in the estimate. The full observability model performs surprisingly well under misspecification. The absolute bias was greater than 0.03 in only about 10% of the simulations. The partial observability model performs surprisingly poorly. The absolute bias was greater than 0.03 in about 85% of the simulations and greater than 0.55 in 10% of the simulations. The figure shows that the errors are not symmetrically distributed around the true probability, however. This suggests that the partial observability model tends to underestimate the occurrence of the event of interest and can drastically underestimate it. However, overestimation is possible as well, making the inferences even more uncertain.

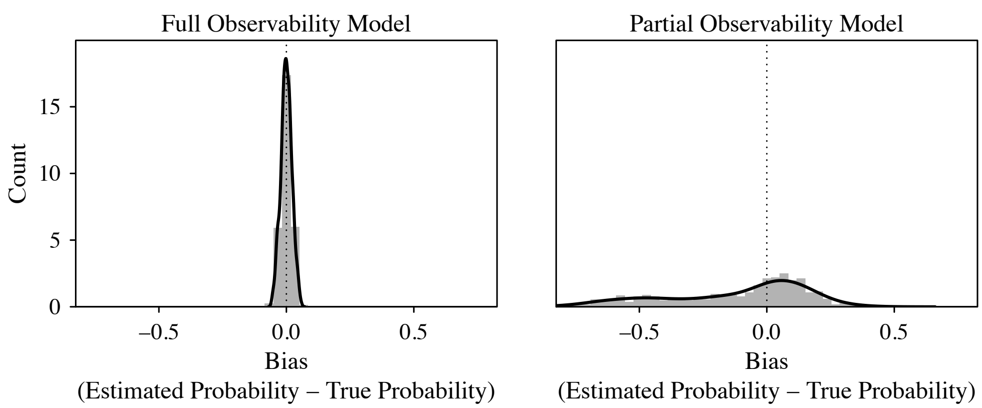

Figure 1: This figure compares the ability of the full and partial observability models to estimate the occurrence probability of the event of interest when the explanatory variable is set to a central value. Notice that the full observability model performs surprisingly well while the partial observability model performs quite poorly.

Next, we examine the ability of the partial observability model to estimate how the probability of the event of interest changes as the key explanatory variable changes from a low value to a high value (0.25 to 0.75 in our case). Figure 2 shows the results. Again, notice that the full observability estimate produces much less bias than the partial observability model. The absolute bias in the estimated effect from the full observability model is greater than 0.03 in only 10% of the simulations. For the partial observability model, the absolute bias is greater than 0.03

in 80% of the simulations and greater than 0.25 in 10% of the simulations. An effect of 0.25 is quite large in political science. This means that the partial observability model is prone to large sample biases — in fact, larger than some of the effects it is being used to estimate.

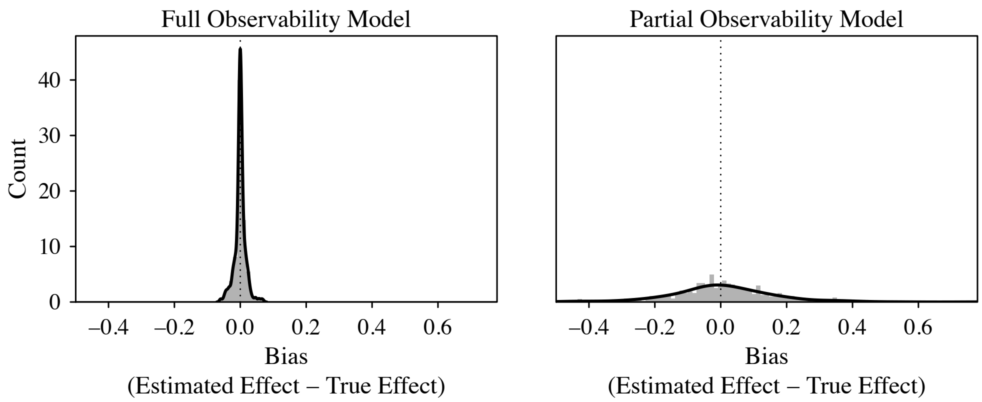

Figure 2: This figure compares the ability of the full and partial observability models to estimate how the occurrence probability of the event of interest changes as the key explanatory variable moves from a low value to a high value. Notice that the full observability model performs surprisingly well while the partial observability model performs quite poorly. Indeed, the partial observability model is prone to biases larger than many of the effects it is used to study.

To get a stronger sense of patterns in the bias, we created scatterplots of the estimates against the true values. Figure 3 shows the estimated probabilities against the true probabilities for both models when x is set at a central value. First, notice the strong correlation between the true probability and the estimated probability from the full observability model (Kendall's τ= 0.93).[10] Second, notice the weak relationship between the true probability and the estimated probability from the partial observability model (Kendall's τ= 0.20). The right panel also re-emphasizes the tendency of partial observability models to underestimate the probability of the event of interest, even when the event is quite likely.

[10] Kendall's tau is comparable to Pearson's r, but is a rank-based statistic and thus more robust to data that do not come from a multivariate normal distribution.

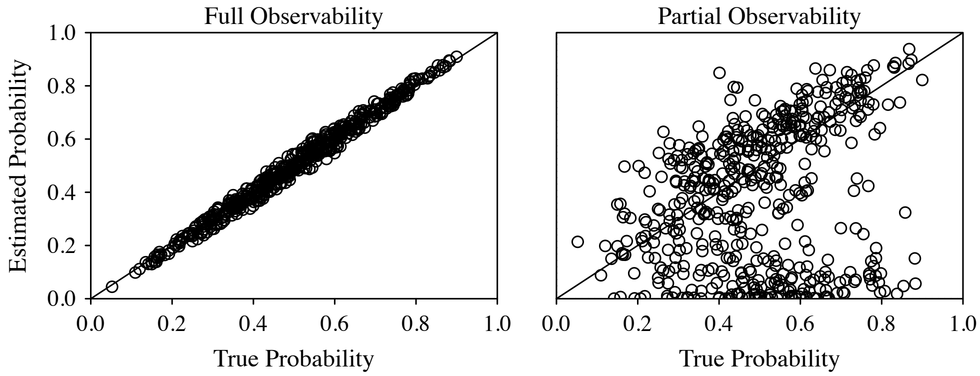

Figure 3: This figure compares the relationship between the true probability and the estimated probability of the event of interest. The solid 45 degree line is $y = x$ and represents a one-to-one correspondence between the true value and the estimate. If a point falls on the line, then the estimate is exactly correct. Notice the strong correlation between the estimated probability and the true probability under full observability and the much weaker correlation under partial observability. Also notice the tendency of partial observability models to underestimate dramatically the probability of an event, even when the event is quite likely.

Figure 4 shows the scatterplots of the estimated and true effects for both models. Notice an even greater correlation between the true effect and the estimated effect under full observability (Kendall's τ= 0.99). The partial observability model does a substantially better job estimating effects than estimating probabilities, but the correlation between the true effect and the partial observability estimate is still relatively weak (Kendall's τ= 0.62). The partial observability model does tend to get the sign of the estimate correct, however. For the partial observability model, only 40 of the 500 (8%) simulations produce a sign error. However, the full observability model produces no sign errors.

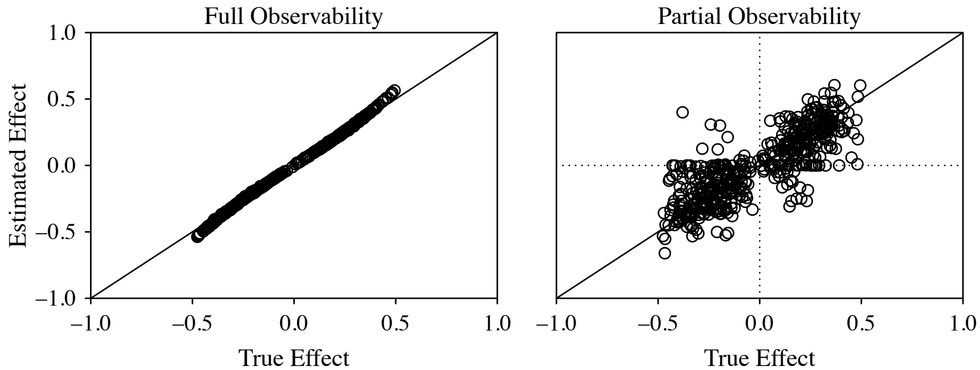

Figure 4: This figure compares the relationship between the true effect and the estimated effect on the event of interest. The solid 45 degree line is y = x and represents a one-to-one correspondence between the true value and the estimate. If a point falls on the line, then the estimate is exactly correct. Notice the nearly perfect correlation between the estimated effect and the true effect under full observability and the much weaker correlation under partial observability. However, notice that the partial observability model does tend to get the sign of the effect correct.

These simulations illustrate what we think is an under-appreciated feature of partial observability models — they are quite sensitive to seemingly innocuous specification errors. Standard errors are often thought to reflect the uncertainty in an estimate. However, standard errors only report the uncertainty due to sampling variation. There are other sources of uncertainty as well, including specification uncertainty. If a researcher is uncertain about the model specification, then the standard errors from a default model are too small. In the case of full observability, our simulations show that the additional uncertainty is perhaps negligible. But in the case of partial observability, the additional uncertainty is quite meaningful. The biases introduced from specification error can be quite large and are rarely negligible. But these results are based on simulations of a hypothetical data generating process (though we think they are realistic and less messy than actual research problems), so we now turn to a real world example that helps to illustrate the problem.

Empirical Example: The Effect of Education on Voter Turnout

Much like the unobservable concept of relevancy lies in the background of studies of conflict, so do misreports lie in the background of survey data. For example, scholars have debated for years the severity of the problem of "over-reporting" voter turnout and how it should be resolved. While labor-intensive efforts can provide validated turnout data,[11] the self-report remains the most common way to gather information about who turns out to vote (and who does not). The partial observability model seems to offer some hope of "modeling away" the problem of misreports in self-reported turnout data (see Beger et al. 2011). Fortuitously, the availability of data based on efforts to validate self-reports for several years of the National Election Studies (NES) offers an unusual opportunity to evaluate the performance of the partial observability model against an approximately correct inference (based on the validated data).

Figure 5 shows the proposed process that leads to self-reported turnout data. Citizens first decide whether to vote. Next, if they abstain from voting, they must choose whether to misreport their decision to the survey interviewer post election. Thus, self-reports of voting result from two related but distinct processes. Respondents self-report voting if *either* (1) they actually voted *or* (2) they choose to misreport their abstention.

[11] Thankfully, the increasing sophistication of electronic databases has reduced the actual financial cost of validation efforts (see Ansolabehere and Hersh 2012).

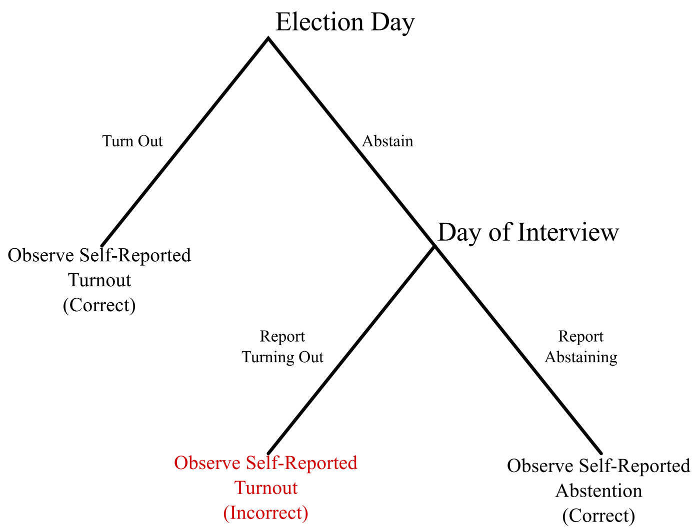

Figure 5: This figure shows the process that we believe generates self-reported turnout data. Notice that the observations from this process are imperfect. Respondents sometimes inaccurately report having turned out to vote. However, a partial observability model might help identify these cases and purge the bias that they introduce into the estimates.

For our primary analysis, we rely on the 1988 NES, the most recent study of a presidential election for which the NES undertook a voter validation effort. The NES is an in-person, cluster-sample survey, with validation data based on NES staff members' in-person examination of local hard-copy registration and voting records. Beginning in 1984, members of the NES staff intensified their searches for missing records of self-reported voters (Cassel 2003), thus these 1988 validated data are among the most accurate in the NES series.[12] [13]

[12] We also repeated our analyses using the 1984, 1986, and 1990 NES data. Although some specific substantive findings vary year to year, the results consistently reinforce our claim that the partial observability model does not perform well under model misspecification.

[13] Although we treat the validated data as fully observed, as Berent, Krosnick, and Lupia (2011) discuss, validation efforts are not without some difficulties. We undertook two additional sets of analyses that took into account the quality of voter validation. First, we considered the sub-sample of respondents whose interviewer assessed as excellent the organization, accessibility, and accuracy of records in the elections office (i.e., the interviewer was confident that if the respondent was registered, his/her record would be found) (see Abramson and Claggett 1992). Second, we also considered the sub-sample of respondents for whom the NES data indicate that there were exact record matches on name, address, and year and month of birth. Each of these sub-samples should be associated with especially high-quality validation efforts. The patterns of results for these sub-samples were consistent with what we present.

[13] We also assessed the validated data available for the 2008 Cooperative Congressional Election Study (CCES). These are the product of the first validation effort conducted in conjunction with a national survey since the NES discontinued its efforts following the 1990 election. However, these data do not contain variables comparable to those available in the NES in terms of providing identification leverage for our partial observability model. Most importantly, the bulk of CCES post-election surveys are completed within a much more compressed time frame relative to Election Day than are the 1988 NES post-election interviews.

Our goal is to study the effect of education on actually turning out. Thus, our main equation models the probability of turning out to vote, and the nuisance equation models the probability of misreporting.[14] We model turnout as a function of several explanatory variables, specifically education, race, and age. We model misreports as a function of these same variables, plus the number of days between the election and the (post-election) interview.[15]

We now turn to the empirical analysis. Our goal is two-fold. First, we assess the differences in the inferences from the typical approach in the literature (i.e., using self-reports as a measure of actually turning out) and the inferences from the partial observability model. Second, we compare the inferences from both approaches to the approximately correct inferences made from validated turnout data.

Constructing an Identifying Variable

To identify the model, we need a variable that predicts cases of misreport and that is exogenous to the decision to turn out. As Przeworski and Vreeland (2002, p. 103) note:

> "As long as at least one variable is assigned to [one equation] that is not assigned to [the other equation], the model can, in principle, be identified. Hence, a strong prior belief that a single variable belongs [in one equation] and not [in the other] is required to identify the parameters."

For identification, we use a variable indicating the number of days after the election that the interview was conducted. Previous research indicates that memory failure likely plays a role in misreports. The most direct evidence shows that the rate of misreporting appears to increase with the time between Election Day and the interview (Belli, Traugott, and Beckman 2001; Stocke 2007; Stocke and Stark 2007; Selb and Munzert 2011). This memory model suggests that

[14] Reassuringly, in the 1988 NES, not a single person who self-reported as not having voted was recorded as having voted in the validated data.
[15] The details of the variables are included in the Appendix

as the number of days increases, over-reporting should increase as well.[16] However, memory failure (which occurs after Election Day) should not affect the decision to vote. Unlike most work relying on partial observability, we can theorize about *and empirically verify* the relationship between our identifying variable and the unobserved dichotomous outcomes. Most other work must rely on theory alone.[17] We indeed find that the date of the interview is (1) positively related to the likelihood of misreporting and (2) only weakly related to the likelihood of actually turning out to vote. When specified in a logistic regression model of misreport, the log of the number of days demonstrates a statistically significant effect ($p = 0.02$) on misreport.[18] However, when included in a model of validated turnout, the log of the number of days is statistically insignificant ($p = 0.35$), and the confidence interval contains only small effects, suggesting that our choice of identifying variable is appropriate. With both a theoretical argument and empirical evidence supporting our identification strategy, the ability of the partial observability to "model away" effectively misreports seems even more promising.

Estimating the Effect of Education on Turning Out

As a test case, we carefully examine the model's ability to estimate accurately the effect of education on turning out. Education is an appealing test case for at least two reasons. First, the relationship between education and political participation is among the most-studied in political science.[19] Second, estimating the effect of education allows us to specify relatively few control

[16] Social desirability also likely factors into the process, since misreports almost always take the form of over-reports (as contrasted with underreports) (see Belli, Traugott, and Beckman 2001; Bernstein, Chadha, and Montjoy 2001; Stocke 2007; Stocke and Stark 2007; Ansolabehere and Hersh 2012).
[17] The existing applications of which we are aware rely on theory alone. Although Stone (2008, fn. 45) reports that he does not find support for Vreeland's (2003) identification strategy regarding Vreeland's partial observability model of IMF agreements, he does not clarify the basis on which he reaches this conclusion.
[18] This model includes all of the key variables that we use in our analysis — years of education, an indicator for African-American respondents, age in years, and age in years squared, along with the number of days between the election and the interview.
[19] A twist has recently appeared with a veritable explosion of new research, relying on a range of methodological approaches (e.g., matching, instrumental variables, natural experiments, and field experiments) to reassess whether

variables, since potentially intervening post-treatment variables, such as political efficacy and strength of partisan attachment, should be excluded (see Gelman and Hill 2007).[20] The simpler model required by focusing on the effect of education reduces the risk of specification error, making the outlook even more hopeful.[21]

To understand the performance of the partial observability model, we estimate three different models. First, we estimate a full observability logit model using the validated data. Since the main event of interest is observed, this model represents the approximate truth. The black points and lines in Figure 6 present the estimates and 90% confidence intervals from this model. For comparison, we also used the validated turnout data to identify misreports and assessed this outcome in a second full observability model of the nuisance outcome of misreporting. This is not of primary interest, but we include it in the right panel of Figure 6 for comparison. Second, we estimate the partial observability logit model using the self-reported turnout data. If the model is performing well, it should help recover the estimates from the full observability model of the validated data. These estimates and 90% confidence intervals are presented in blue in Figure 6. Finally, we estimate a "naive" model in which we ignore the misreports and treat the self-reported data as though they measure actual turnout. This has been the approach of most previous researchers.[22] We include these estimates to get a sense of whether the partial observability model at least reduces the bias introduced by misreporting.

---

education has a *causal* role (e.g., Dee 2004; Kam and Palmer 2008, 2011; Sondheimer and Green 2010; Henderson and Chatfield, 2011; Mayer 2011; Berinksy and Lenz 2011). We do not weigh in on this unfolding debate in the current study.

[20] Although the conventional approach of most students of voter turnout is to specify both demographic and attitudinal variables in a multivariate model, Gelman and Hill (2007, pp. 190-94) warn against making inferences based on models that specify intervening or mediating variables, which attitudinal measures clearly are in this set-up. The concern is "nonignorability—systematic differences between groups defined conditional on the post-treatment intermediate outcome" (Gelman and Hill, p. 193).

[21] Though less substantively important, the simpler model also helps us to avoid separation and convergence issues that typically plague partial observability models.

[22] Obvious exceptions are those studies that have relied on validated turnout data.

Figure 6 shows that the naive model (ignoring misreports) overestimates the coefficient for education by about 33%. Although not large, this bias is substantively meaningful. The partial observability model working well would reduce, if not eliminate, this upward bias. Instead, invoking the model triples the bias, overestimating the effect by 100%. That is, the partial observability model, rather than "modeling away" the misreports, exacerbates their effect on the estimates. Readers might notice similarly troubling differences across the other coefficients, but we are hesitant to directly interpret these coefficients because education likely mediates the other variables.

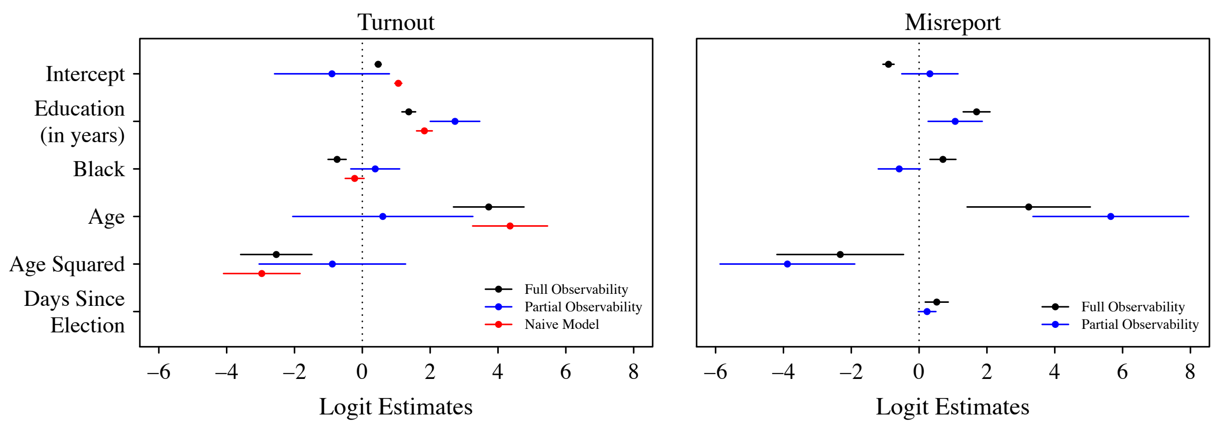

Figure 6: Estimates of full and partial observability models of turning out to vote and misreporting turning out. The full observability models uses validated vote data to estimate direct models of turnout and misreport. These estimates and 90% confidence intervals represent the approximate truth. The partial observability model uses self-reported turnout data and attempts to parse out the effects of the explanatory variables on the partially observed outcomes of turning out and misreporting using the identifying variable of number of days between the election and the interview. The naive model ignores misreports by treating self-reported turnout as actual turnout. The models are designed to estimate the effect of education on turnout. Notice that the partial observability model exacerbates, rather than reduces, the bias due to misreporting in the coefficient for education.

However, analysts might care more about the ability of the partial observability model to recover the predicted probabilities, marginal effects, or first differences rather than model coefficients. Figure 7 presents the estimated probability of voting as education varies. For these results, we set all other explanatory variables to their sample medians. The naive approach of ignoring misreports leads to a *slight* underestimation of the probability of voting for low values of education and a *slight* overestimation for high values of education. But notice the estimates from the partial observability model, which severely underestimate the probability of voting for those with low levels of education. Only for those with the highest levels of education do the partial observability estimates begin to approach the estimates using the validated vote data.

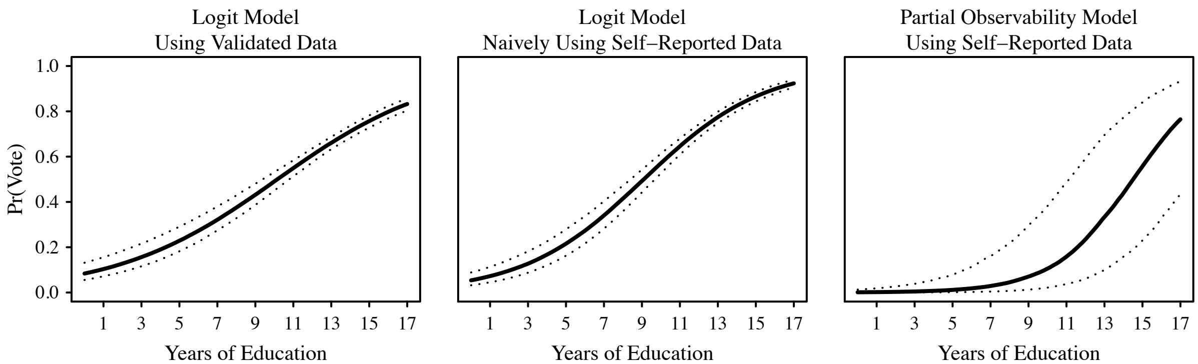

Figure 7: This figure shows how the estimated probability of voting varies across education for three models. The left panel shows the full observability model (using validated voting data), which we take to represent the approximate truth. The middle panel shows the naive model that ignores misreports by treating self-reported data as actual turnout data. Finally, the right panel shows the estimates from the partial observability model using self-reported data. Notice that ignoring the misreports produces only small biases, while attempting to "model away" the misreports produces much larger biases.

Of perhaps greater importance is the estimated effect of education on turnout. Figure 8 presents the marginal effect of education as education varies across the three models. Notice first that the naive model (ignoring misreports) slightly overestimates the marginal effect of education, especially among those with 12 years of education (a high school diploma) or less. The partial observability model, on the other hand, dramatically underestimates the marginal effect of education among those with 10 years of education or less, while dramatically overestimating the marginal effect for those with more than 10 years of education, which is most of the sample (83%).

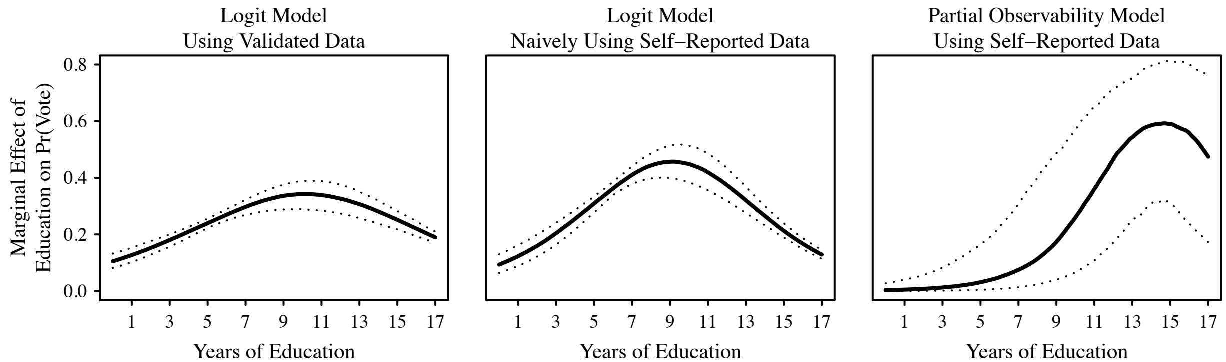

Figure 8: This figure shows the marginal effect of education on the probability of voting as education varies for three different models. The logit model using the validated data provides the approximate truth. The logit model naively using self-reported data shows the small biases introduced by ignoring misreports.

But perhaps the marginal effect is not the quantity of interest. Perhaps the interesting effect is the estimated change in the probability of voting as one moves from a high school diploma (12 years of education) to a college degree (16 years of education). This change may be the most substantively interesting (see Kam and Palmer 2008), and these represent the two most common education levels in the data (about 35% and 12%, respectively). Figure 9 shows the estimated first-differences. Notice first that the full observability model using the validated turnout data suggests that increasing a respondent's years of education from 12 to 16 increases her chance of voting by about 19 percentage points. The naive model estimates this particular effect quite well, suggesting a slightly smaller increase of about 18 percentage points. The partial observability model, however, drastically overestimates the effect, suggesting an effect of about 38 percentage points. This means that the partial observability model overestimates the effect by about 100%.

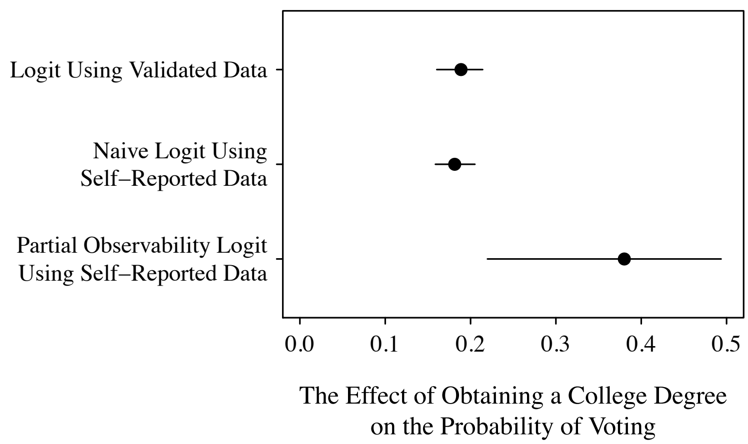

Figure 9: This figure shows the estimated change in the probability of voting from three models. The full observability model using the validated data represents the approximate truth. The naive logit uses ignores misreports and treats the self-reports as actual turnout. The partial observability model attempts to "model away" any bias introduced by misreports. Instead of reducing the bias, however, it exacerbates bias.

This empirical illustration shows that one must be quite careful about using a partial observability model to parse out the separate effects of a single explanatory variable on two jointly observed outcomes. Indeed, if not for our ability to check our inferences using the validated turnout data, we likely would have concluded that misreports lead researchers to underestimate dramatically the effect of education on turnout. Most applications of partial observability models do not enable checking inferences in a similar manner. Thus, these analyses depend on the assumption that the model specification is correct or nearly so. Indeed, our simulations consider a situation in which the analyst knows quite a bit about a simple data generating process. Even in this manageable situation, however, the inferences are highly sensitive to model specification — much more sensitive than a full observability logit model of the same process.

Conclusion

We began this study optimistic about using partial-observability to model measurement error in binary outcome variables (see especially Beger et al. 2011), particularly in self-reported turnout data. We wish that we had discovered evidence to begin to make a compelling case that the partial observability model is a viable option for addressing the over-report of voter turnout phenomenon among survey respondents (when validated are not available) – alas, we have not. Indeed, both our simulations and empirical example suggest that partial observability model might lead to biases larger than the measurement error they are intended to correct.

The fact that the traditional approach of simply ignoring misreports does not lead to especially misleading inferences for the effect that we examine may provide some comfort to consumers of self-reported NES turnout data; however, we are not suggesting that our results regarding the "status quo" approach should dissuade efforts to gather higher-quality turnout data (see Presser 1990; Belli, Traugott, Young, and McGonagle 1999; Belli, Moore, and Van Hoewyk 2006; Ansolabehere and Hersh 2012; Hanmer, Banks, and White 2014). On the contrary, higher quality data offer the most compelling solution to the problems that we encounter. Be neither do we want to suggest completely abandoning complex modeling approaches (e.g., Katz and Katz 2010).[23] However, we do suggest that researcher interpret the results with care and subject the findings to a wide range of robustness checks, such as alternative measurement of key variables and alternative model specifications.

More generally, our results give us pause regarding the inferences from existing studies that employ a partial observability model. Our understanding of electoral participation is comparable (and perhaps even superior) to that of many areas of social science, and our specification, although purposefully parsimonious, strikes us as theoretically-grounded. Furthermore, unlike other researchers who have employed a partial observability model, we are able to invoke validated data to assess our specifications using a full observability model. In

[23] Moreover, Ansolabehere and Hersh's (2011, 2012) recent findings based on comparing the results of voter turnout models that rely on validated data to those of models that rely on self-reports are not so sanguine. They determine that standard predictors of voter turnout, such as demographics and measures of partisanship and political engagement, explain only a third to half as much about participation in models that assess validated data.

addition, our rationale regarding the selection of our identifying variable is not unlike that found in existing efforts that employ a partial observability model—except we are actually able to confirm empirically that our identifying variable is a significant predictor of misreport and not of actually voting. Indeed, everything seems to work in our favor for a successful application of the partial observability model. Despite this, our simulations and empirical illustration show that inferences from a partial observability model depend on an (unrealistically) accurate model specification. The potential price of our exercise is learning that the partial observability model comes up wanting — and that has happened. However, the risks of proceeding blindly and not acknowledging the potential problems of this model are drawing overconfident inferences and reaching erroneous conclusions.

References

Abramson, Paul R., and William Claggett. 1984. "Race-related Differences in Self-Reported and Validated Turnout." Journal of Politics 46:719-738.

Abramson, Paul R., and William Claggett. 1992. "The Quality of Record Keeping and Racial Differences in Validated Turnout." *Journal of Politics* 54:871-880.

Ansolabehere, Stephen, and Eitan Hersh. July 14, 2010. "The Quality of Voter Registration Records: A State-by-State Analysis." Working paper published by the Institute for Quantitative Social Science at Harvard University and the Caltech/MIT Voting Technology Project.

Ansolabehere, Stephen, and Eitan Hersh. November 7, 2011. "Misreporting, Sample Selection, and Political Participation." Working paper.

Ansolabehere, Stephen, and Eitan Hersh. 2012. "Validation: What Big Data Reveal About Survey Misreporting and the Real Electorate." *Political Analysis* 20:437-459.

Beger, Andreas, Jacqueline H.R. DeMeritt, Wonjae Hwang, and Will H. Moore. February 28, 2011. "The Split Population Logit (SPopLogit): Modeling Measurement Bias in Binary Data." Working paper.

Belli, Robert F., Michael W. Traugott, and Steven J. Rosenstone. 1994. "Reducing Over-Reporting of Voter Turnout: An Experiment Using a Source Monitoring Framework." NES Technical Reports Number 35.

Belli, Robert F., Michael W. Traugott, Margaret Young, and Katherine A. McGonagle. 1999. "Reducing Vote Over-reporting in Surveys: Social Desirability , Memory Failure, and Source Monitoring." *Public Opinion Quarterly* 63:90-108.

Belli, Robert F., Sean E. Moore, John VanHoewyk. 2006. "An Experimental Comparison of Question Forms Used to Reduce Vote Overreporting." *Electoral Studies*:751-759.

Berent, Matthew K., Jon A. Krosnick, and Arthur Lupia. August 2011. "The Quality of Government Records and 'Over-estimation' of Registration and Turnout in Surveys: Lessons from the 2008 ANES Panel Study's Registration and Validation Exercises." Working paper no. nes012554. Ann Arbor, MI and Palo Alto, CA: American National Election Studies.

Berinsky, Adam J., and Gabriel S. Lenz. 2011. "Education and Political Participation: Exploring the Causal Link." *Political Behavior* 33:357-373.

Bernstein, Robert, Anita Chadha, and Robert Monjoy. 2001. "Over-reporting Voting: Why It

Happens and Why It Matters." *Public Opinion Quarterly* 65:22-44.

Berry, William, Jacqueline H. R. Demeritt, and Justin Esarey. 2014. "Bias and Overconfidence in Parametric Models of Interactive Processes." Forthcoming at *American Journal of Political Science*.

Braumoeller, Bear F. 2003. "Causal Complexity and the Study of Politics." *Political Analysis* 11:209-233.

Cassel, Carol A. 2002. "Hispanic Turnout: Estimates from Validated Voting Data." *Political Research Quarterly* 55:391-408.

Cassel, Carol A. 2003. "Over-reporting and Electoral Participation Research." *American Politics Research* 31:81-92.

Cassel, Carol A. 2004. "Voting Records and Validated Voting Studies." *Public Opinion Quarterly* 68:102-108.

Clausen, Aage R. 1968. "Response Validity: Vote Report." *Public Opinion Quarterly* 32:588-606.

Dee, Thomas S. 2004. "Are There Civic Returns to Education?" *Journal of Public Economics* 88:1697-1720.

Deufel, Benjamin J., and Orit Kedar. 2010. "Race and Turnout in U.S. elections: Exposing Hidden Effects." *Public Opinion Quarterly* 74:286-318.

Duff, Brian, Michael J. Hanmer, Won-Ho Park, and Ismail K. White. 2007. "Good Excuses: Understanding Who Votes With an Improved Turnout Question." *Public Opinion Quarterly* 71:67-90.

Gelman, Andrew, and Jennifer Hill. 2007. *Data Analysis Using Regression and Multilevel/Hierarchical Models*. Cambridge, NY: Cambridge University Press.

Hanmer, Michael J., Antoine J. Banks, and Ismail K. White. 2014. "Experiments to Reduce the Over-reporting of Voting: A Pipeline to the Truth." *Political Analysis* 22:130-141.

Henderson, John, and Sara Chatfield. 2011. "Who Matches? Propensity Scores and Bias in the Causal Effects of Education on Participation." *Journal of Politics* 73:646-658.

Hill, Kim Quaile, and Patricia A. Hurley. 1982. "Nonvoters in Voters' Clothing: The Impact of Voting Behavior Misreporting on Voting Behavior Research." *Social Science Quarterly* 65:199-206.

Holbrook, Allyson L., and Jon A. Krosnick. 2010. "Social Desirability Bias in Voter Turnout Reports: Tests Using the Item Count Technique." *Public Opinion Quarterly* 74:37-67.

Kam, Cindy, and Carl Palmer. 2008. “Reconsidering the Effects of Education on Political Participation.” *Journal of Politics* 70:612-631.

Kam, Cindy D., and Carl P. Palmer. 2011. “Rejoinder: Reinvestigating the Causal Relationship between Higher Education and Political Participation.” *Journal of Politics* 73:659-663.

Katosh, John P., and Michael W. Traugott. 1981. “The Consequences of Validated and Self-Reported Voting Measures.” *Public Opinion Quarterly* 45:519-535.

Katz, Jonathan N., and Gabriel Katz. 2010. “Correcting for Survey Misreports Using Auxiliary Information with an Application to Estimating Turnout.” *American Journal of Political Science* 54:815-835.

Malhotra, Neil, and Jon A. Krosnick. 2007. “The Effect of Survey Mode and Sampling on Inferences About Political Attitudes and Behavior: Comparing the 2000 and 2004 ANES to Internet Surveys With Nonprobability Samples.” *Political Analysis* 15:286-323.

Mayer, Alex. 2011. “Does Education Increase Participation.” *Journal of Politics* 73:633-645.

Poirier, Dale J. 1980. “Partial Observability in Bivariate Probit Models.” *Journal of Econometrics* 12:209-217.

Presser, Stanley. 1990. “Can Changes in Context Reduce Vote Over-reporting in Surveys?” *Public Opinion Quarterly* 54:586-593.

Presser, Stanley, Michael W. Traugott, and Santa Traugott. 1990. “Vote ‘Over’ Reporting in Surveys: The Records or the Respondents?” Technical Report no. 39. Ann Arbor, MI: National Election Studies.

Przeworski, Adam, and James Raymond Vreeland. 2000. “The Effect of IMF Programs on Economic Growth.” *Journal of Development Economics* 62:385-421.

Przeworski, Adam, and James Raymond Vreeland. 2002. “A Statistical Model of Bilateral Cooperation.” *Political Analysis* 10:101-112.

Selb, Peter, and Simon Munzert. March 16, 2011. “Voter Overrepresentation, Vote Misreporting, and Turnout Bias in Postelection Surveys.” Working paper.

Sigelman, Lee. 1982. “The Nonvoting Voter in Voting Research.” *American Journal of Political Science* 26:47-56.

Silver, Brian D., Paul R. Abramson, and Barbara A. Anderson. 1986. “The Presence of Others and Over-reporting of Voting in American National Elections.” *Public Opinion Quarterly* 50:228-239.

Silver, Brian D., Barbara A. Anderson, and Paul R. Abramson. 1986. "Who Over-reports Voting?" *American Political Science Review* 80:613-624.

Sondheimer, Rachel Milstein, and Donald P. Green. 2010. "Using Experiments to Estimate the Effects of Education on Voter Turnout." *American Journal of Political Science* 54:174-189.

Stocke, Volker. 2007. "Reponse Privacy and Elapsed Time Since Election Day as Determinants for Vote Over-reporting." *International Journal of Public Opinion Research* 19:237-246.

Stocke, Volker, and Tobias Stark. 2007. "Political Involvement and Memory Failure as Interdependent Determinants of Vote Over-reporting." *Applied Cognitive Psychology* 21:239-257.

Stone, Randall W. 2008. "The Scope of IMF Conditionality." *International Organization* 62:589-620.

Traugott, Michael W., and John P. Katosh. 1979. "Response Validity in Surveys of Voting Behavior." *Public Opinion Quarterly* 43:359-377.

Traugott, Santa. 1989. "Validating the Self-Reported Vote: 1964-1988." Technical Report no. 34. Ann Arbor, MI: National Election Studies.

Vreeland, James Raymond. 2003. *The IMF and Economic Development*. Cambridge, UK: Cambridge University Press.

Xiang, Jun. 2010. "Relevance as a Latent Variable in Dyadic Analysis of Conflict." *The Journal of Politics* 72:484-498.

Yalch, Richard F. 1976. "Pre-election Interview Effects on Voter Turnout." *Public Opinion Quarterly* 40:331-336.

Online Appendix I: Data

The primary NES data used in this analysis come from the Third ICSPR Edition (January 2000) release of the American National Election Study, 1988: Pre- and Post-Election Survey (ICPSR 9196).[24] The dependent variables in turnout equations are TURNOUT SELF REPORT (v880756, recoded 5=0) and TURNOUT VALIDATED (v881147, recoded as 11=1; 21, 22, 24, 31, and 32=0; and 0, 12, 13, 23, and 33=missing); our coding choices regarding TURNOUT VALIDATED are based on Cassel (2004). The independent variables in our turnout equations are EDUCATION (v880419), BLACK (v880412, recoded as 2=1 and 1, 3, 4, and 7=0), and AGE (v880417). Additionally, the variable DAYS SINCE ELECTION (based on date of post-election variable, recoded as ((v880063*100)+(v880064)) appears in the misreport equation.

Descriptive Statistics	Obs	Mean	Std. dev.	Min.	Max.
TURNOUT SELF REPORT	1773	.697		0	1
TURNOUT VALIDATED	1708	.598		0	1
EDUCATION	2035	12.54	2.88	0	17
BLACK	2034	.132		0	1
AGE	2037	45.01	17.74	17	99
DAYS SINCE ELECTION	1773	16.43	12.84	1	65

[24] Miller, Warren E., and the National Election Studies. AMERICAN NATIONAL ELECTION STUDY, 1988: PRE- AND POST-ELECTION SURVEY [Computer File]. Conducted by University of Michigan, Center for Political Studies. 3rd ICPSR ed. Ann Arbor, MI: Inter-university Consortium for Political and Social Research [producer and distributor], 2000.

Online Appendix II: Estimation

To estimate the parameters, we derive the log-likelihood functions of the data $\boldsymbol{y}$ given the parameter vector $\boldsymbol{\theta} = \begin{bmatrix}\boldsymbol{\beta} \\ \boldsymbol{\gamma}\end{bmatrix}$. To set up this model in a choice-theoretical fashion, we assume that each individual *i* derives a utility U_i^{vote} from actually voting and a possibly different utility $U^{report}|(U^{vote} < 0)$ from reporting voting conditional on abstaining. If either of these utilities is positive, then the individual reports voting. If U_i^{vote} is positive, then the individual actually votes. If U_i^{vote} is negative, then the individual abstains but reports voting if $U^{report}|(U^{vote} < 0)$ is positive and reports abstaining if $U^{report}|(U^{vote} < 0)$ is negative. To compute the likelihood, note that $L(\boldsymbol{\theta}|\boldsymbol{y}) = \prod_{i=1}^{n} P(y_i = 1)$, where $P(y_i = 1) = P(U_i^{vote} > 0 \text{ or } U^{report}|(U^{vote} < 0) > 0)$. Since there is no overlap between these two events, this is simply the sum of the probabilities of the two events, so that $P(y_i = 1) = P(U_i^{vote} > 0) + \mathrm{P}(U^{report}|(U^{vote} < 0) > 0)P(U^{vote} < 0)$.

The first probability in the sum is easy to compute. We assume that the error in the utility calculations in the vote stage of the model follows the standard logistic distribution, denoted as $f_\varepsilon\left(0, \frac{\pi^2}{3}\right)$. Thus, $P(U_i^{vote} > 0) = F_\varepsilon(\boldsymbol{X}_i\boldsymbol{\beta})$, where $F_\varepsilon(x)$ is the cumulative distribution function (cdf) of f_ε, so that $F_\varepsilon(x) = \int_{-\infty}^{x} f_\varepsilon dt$.

The second term of the sum requires a little more care. We assume that the random components in the utilities are uncorrelated. The product $P(U^{report}|(U^{vote} < 0) > 0)P(U_i^{vote} < 0)$ is of course simply the product of the two probabilities, which are known from the model. In this case, $P(U^{report}|(U^{vote} < 0) > 0)P(U_i^{vote} < 0) = F_\varepsilon(\boldsymbol{Z}_i\boldsymbol{\gamma})[1 - F_\varepsilon(\boldsymbol{X}_i\boldsymbol{\beta})]$. This yields: $P(y_i = 1) = F_\varepsilon(\boldsymbol{X}_i\boldsymbol{\beta}) + F_\varepsilon(\boldsymbol{Z}_i\boldsymbol{\gamma})[1 - F_\varepsilon(\boldsymbol{X}_i\boldsymbol{\beta})]$, the right side which might be written as $Q_i(\boldsymbol{X}_i\boldsymbol{\beta}, \boldsymbol{Z}_i\boldsymbol{\gamma})$. Thus, we can write the likelihood as $\mathcal{L}(\boldsymbol{y}|\boldsymbol{\theta}, \boldsymbol{X}, \boldsymbol{Z}) = \prod_{i=1}^{n}[Q_i^{y_i} + (1 - Q_i)^{1-y_i}]$. Taking the natural log gives $\ln L(\boldsymbol{y}|\boldsymbol{\theta}, \boldsymbol{X}, \boldsymbol{Z}) = \sum_{i=1}^{n}[y_i \ln Q_i + (1 - y_i)\ln(1 - Q_i)]$.

It is possible to maximize the likelihood using only the likelihood function by computing finite differences to determine step direction and size. However, it is more computationally efficient to rely on the gradient vector $\boldsymbol{g}(\boldsymbol{\theta}) = \frac{\partial \ln L}{\partial \boldsymbol{\theta}} = \left[\frac{\partial \ln L}{\partial \beta_1}, \ldots, \frac{\partial \ln L}{\partial \beta_k}, \frac{\partial \ln L}{\partial \gamma_1}, \ldots, \frac{\partial \ln L}{\partial \gamma_m}\right]'$ to determine step direction and size. The elements of this gradient vector are easily computed:

$$\frac{\partial \ln L}{\partial \beta_j} = \sum_{i=1}^{n}\left[\frac{y_i}{Q_i}\left(f_\varepsilon(\boldsymbol{X}_i\boldsymbol{\beta})X_{ij} - F_\varepsilon(\boldsymbol{Z}_i\boldsymbol{\gamma})f_\varepsilon(\boldsymbol{X}_i\boldsymbol{\beta})X_{ij}\right) + \frac{1 - y_i}{1 - Q_i}\left(f_\varepsilon(\boldsymbol{X}_i\boldsymbol{\beta})X_{ij} - F_\varepsilon(\boldsymbol{Z}_i\boldsymbol{\gamma})f_\varepsilon(\boldsymbol{X}_i\boldsymbol{\beta})X_{ij}\right)\right]$$

$$\frac{\partial \ln L}{\partial \gamma_j} = \sum_{i=1}^{n}\left[\frac{y_i}{Q_i}\left(f_\varepsilon(\boldsymbol{Z}_i\boldsymbol{\gamma})Z_{ij} - F_\varepsilon(\boldsymbol{X}_i\boldsymbol{\beta})f_\varepsilon(\boldsymbol{Z}_i\boldsymbol{\gamma})Z_{ij}\right) + \frac{1 - y_i}{1 - Q_i}\left(f_\varepsilon(\boldsymbol{Z}_i\boldsymbol{\gamma})Z_{ij} - F_\varepsilon(\boldsymbol{X}_i\boldsymbol{\beta})f_\varepsilon(\boldsymbol{Z}_i\boldsymbol{\gamma})Z_{ij}\right)\right]$$